

LIVVkit 2: An extensible land ice verification and validation toolkit for comparing observations and models

Joseph H. Kennedy

Andrew R. Bennett, Katherine J. Evans, Jeremy Fyke, Matthew J. Hoffman, Stephen F. Price, Lauren J. Vargo

Confidence and credibility

Confidence: feeling or belief that one can rely on something

Credibility: the quality of being trusted and believed in

■ Verification and validation (V&V) is a set of confidence building techniques.

■ V&V is a continuous process tandem to, and essential to, development.

■ **V&V is not enough!** Credibility relies on:

- Reproducibility
- Transparency
- Discoverability

LIVVkit is designed to build user and developer confidence and scientific credibility

Verification

Software verification

LIVVkit: verification

The process of determining if the software's implementation accurately represents the developers' specifications. This is an **engineering** problem:

"Did we build what we intended?"

Numerical (algorithm) verification

LIVVkit: numerics

The process of comparing the approximate numerical solution of the model, or parts of the model, against a numerical benchmark (e.g., an analytical solution or a manufactured solution). This is a **math** problem:

"Are we solving the equations correctly?"

Validation

Software validation

LIVVkit: performance

The process of determining how well the software is able to be used for its intended task. In the case of ice-sheet models, especially for those coupled to a global-climate model, *performance* aspects will be the focus of software validation. This is a **design** problem:

"Did we build what the users needed?"

Physical validation

LIVVkit: validation

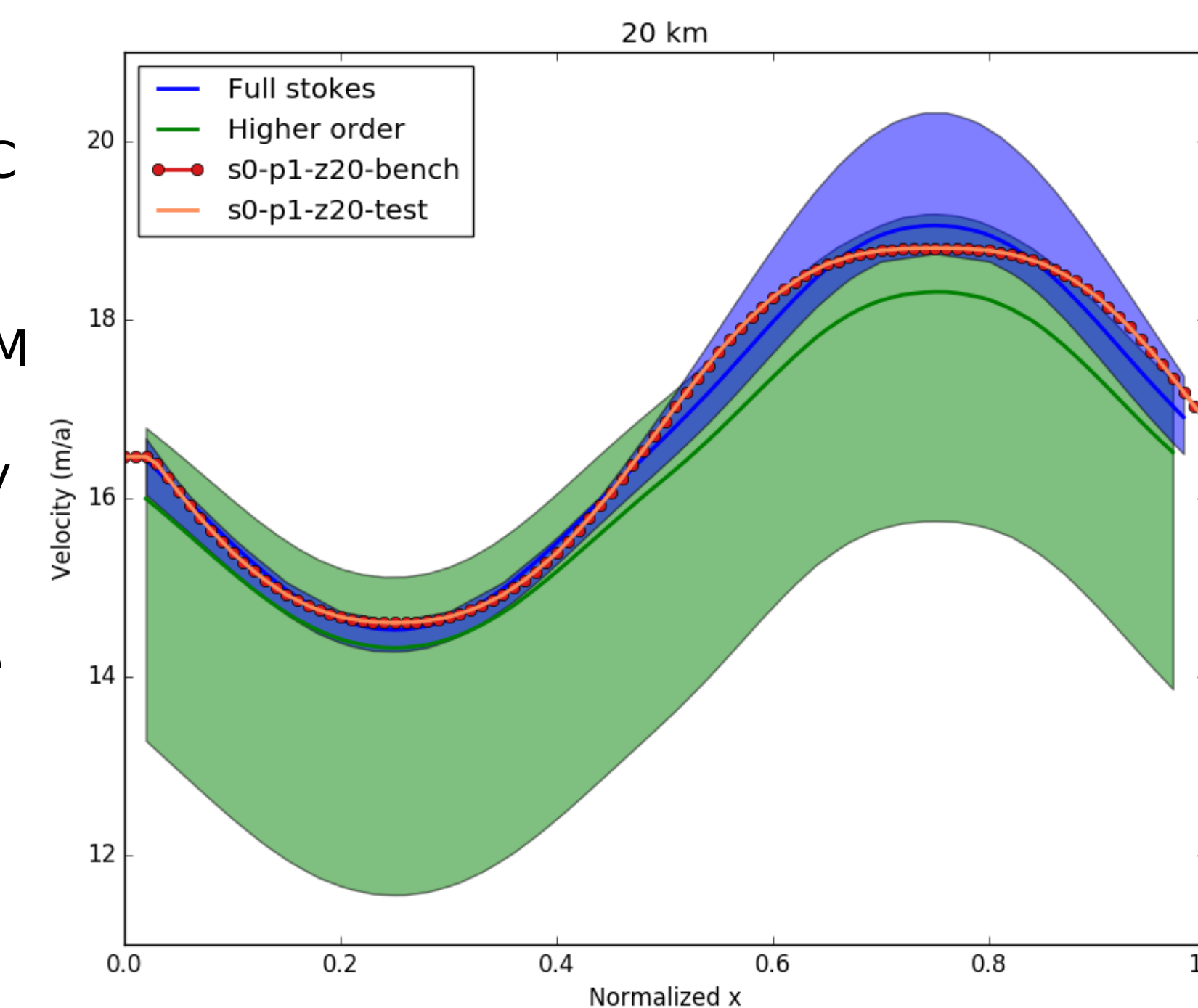
The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. This is a **physics** problem:

"Are we using the right physics?"

Verification analysis

Example verification analysis:

A comparison of CISM ISMIP-HOM C test (orange line) at a 20 km resolution and the CISM benchmark (red dots) with the ISMIP-HOM results. The mean result of the ISMIP-HOM participant models fully solving the Stokes equation is shown by the blue line, while the range of results is indicated by the shaded blue region. Similarly, the green line and shaded region shows the models using a higher order approximation (1st order or higher) to the Stokes equation.



Nightly testing of CISM: <http://jhkennedy.org/nightly/>

Architecture

Python 3 package

- Dependencies satisfied by Anaconda/Miniconda
- Python 2 and python 3 cross-compatibility soon

Command-line interface

- 3 main options

Analysis Scheduler

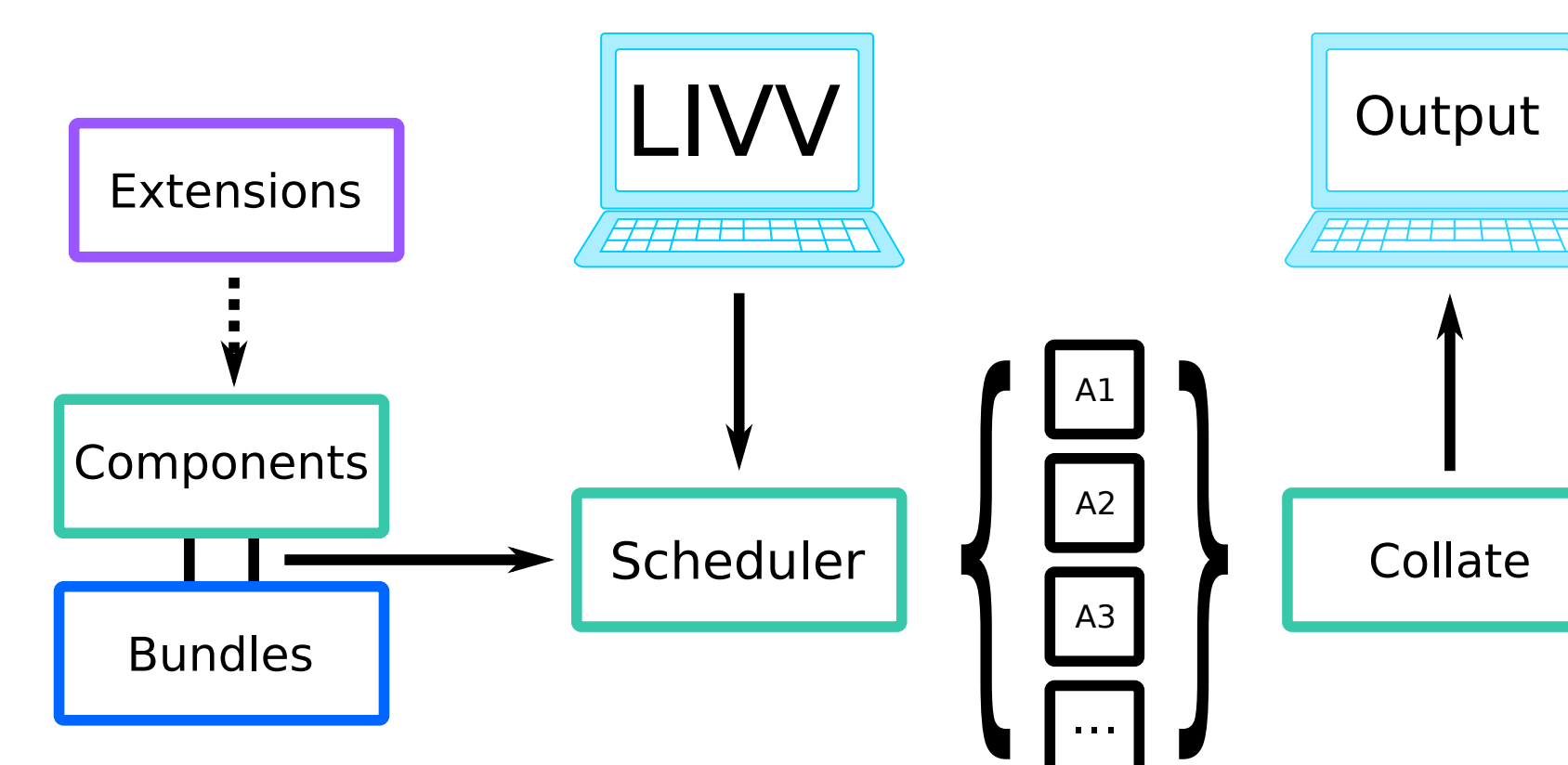
- Analyses run in parallel

Model description 'bundles'

- Encapsulates model specifics
- Easily incorporate new models

Components

- verification, numerics, performance, and validation



Extensions infrastructure

- Custom Analyses
- JSON config files and python code
- Minimal working example provided

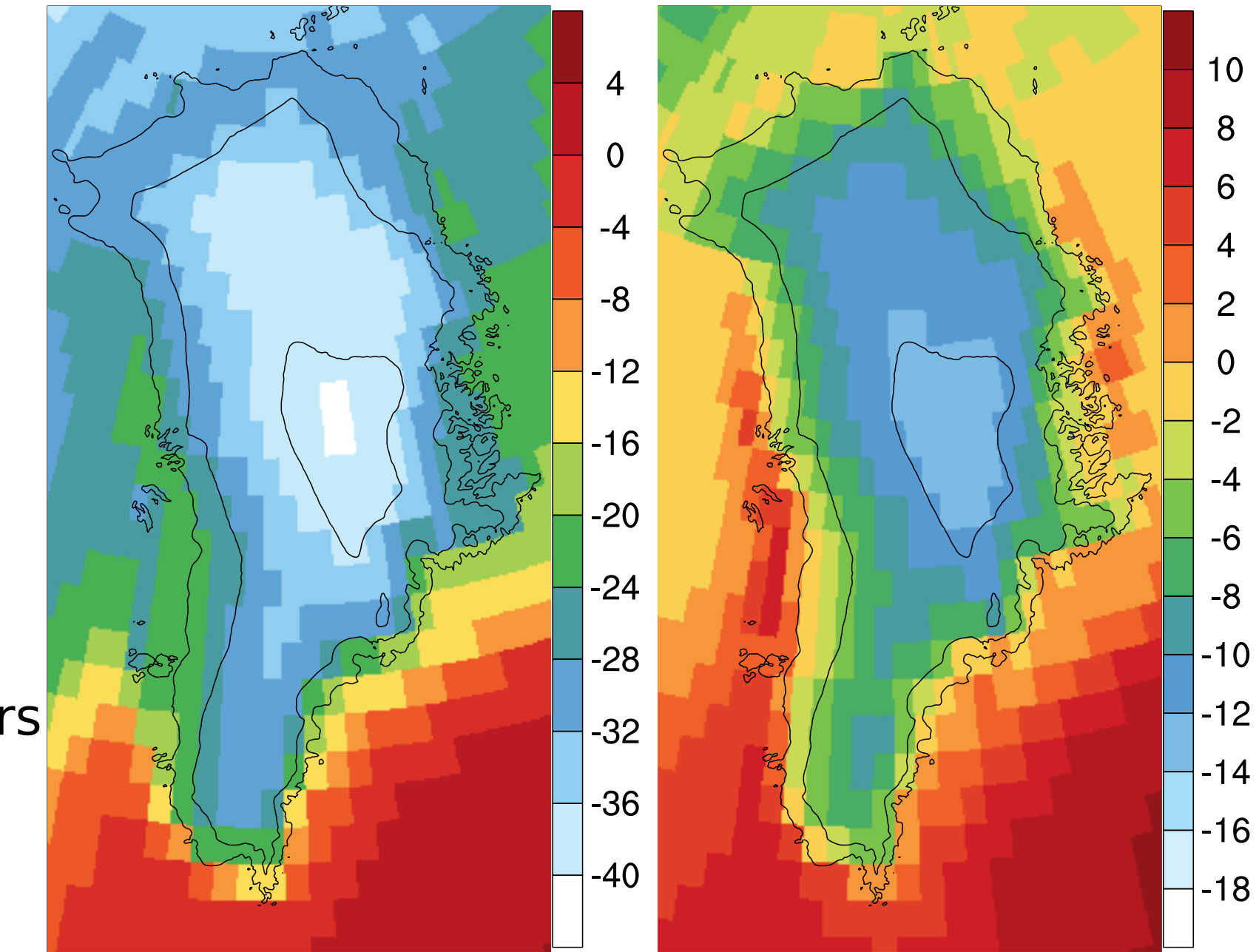
Output website

- Self contained and portable
- HTML and javascript

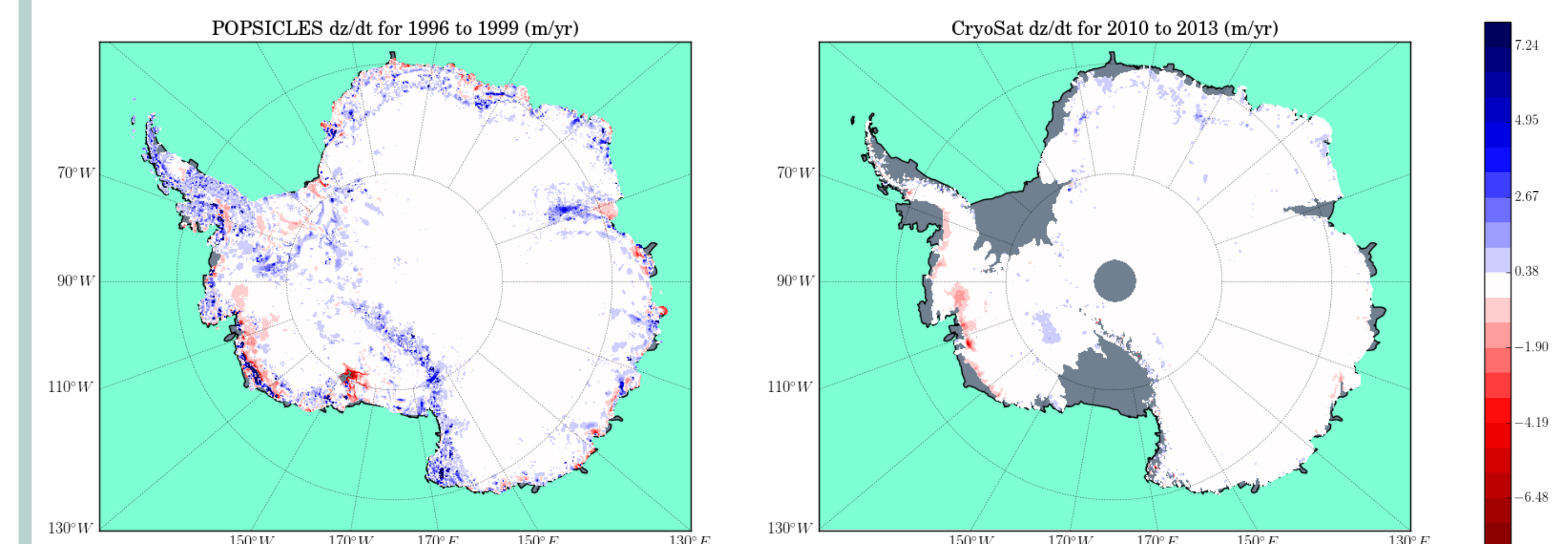
Coupled model diagnostics

Example CESM diagnostic:

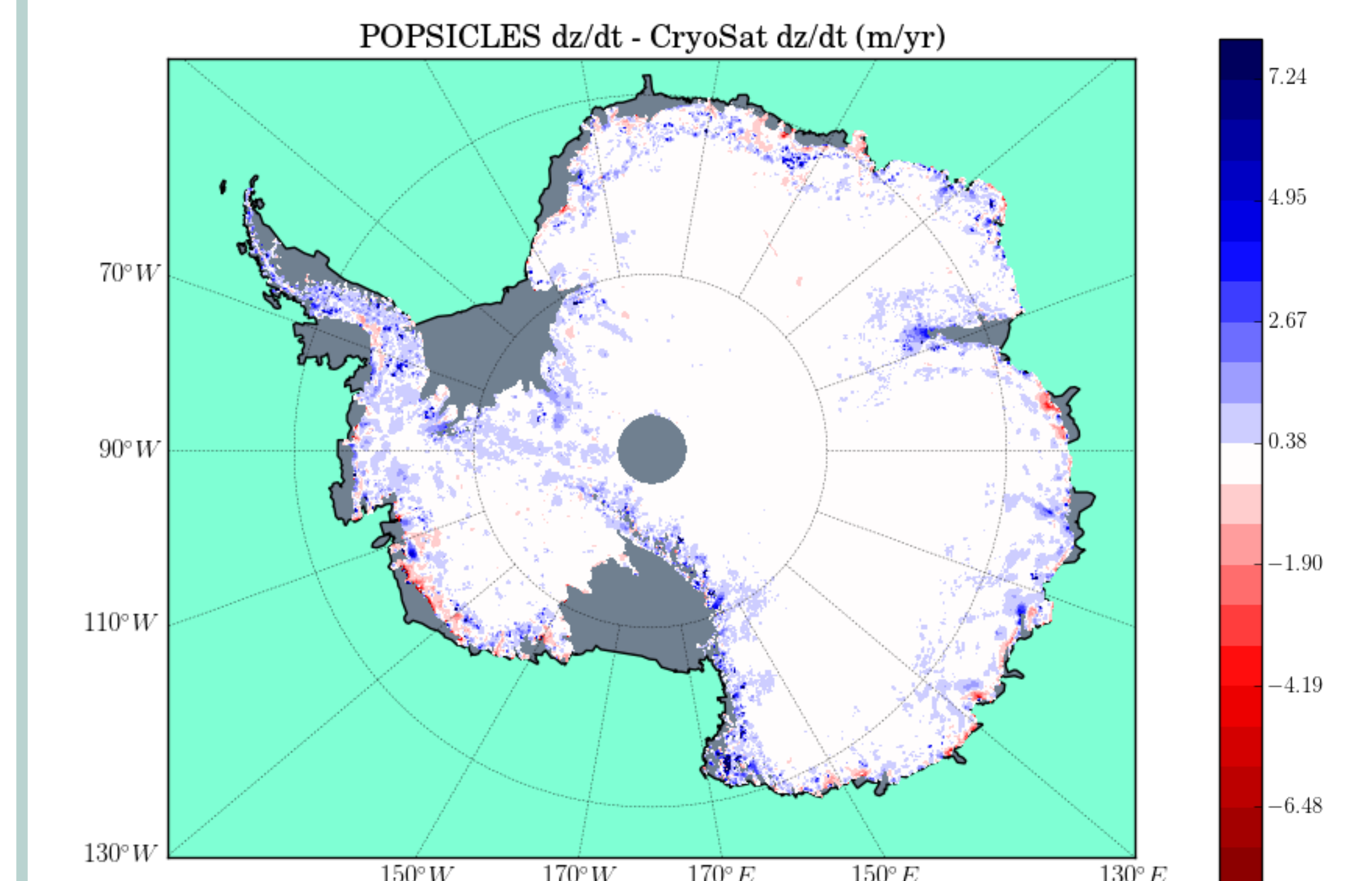
Mean 1960–2005 near-surface temperature (degrees C) as simulated by CESM1.0 as presented in Vizcaino et al. (2013) (left) winter (DJF) and (right) summer (JJA). Over land, values represent the weighted average over the different land surface types. Elevation contours are plotted in black at 0, 1000, 2000, and 3000 m.



Antarctic validation



Above left: Elevation changes (m w.e./yr) for an Antarctic simulation from 1996 to 1998 using POPCICLES. Data was provided by Daniel Martin and Xylar Asay-Davis. Above right: Elevation changes observed over Antarctica for 2010 to 2013 from CryoSat-2 (McMillan et al., 2015). Data was provided by Malcolm McMillan and Andrew Shepherd for validation purposes. Below: The difference between the POPCICLES and CryoSat-2 elevation changes.



Good Design = Good Tools

Usable

- No new technical skills needed
- Simple command-line interface

Portable

- Platform independent; Desktops to HPCs.

Flexible

- Scriptable and importable
- Integrates into many workflows

Extensible

- Quickly accommodates new models and tests

Nimble

- Performs the desired analyses quickly, with little overhead

Informative and sharable

- Results are well described and provide the appropriate context